Biodiversity Data Journal 13: e137540
doi: 10.3897/BDJ.12.e137540

OPEN ACCESS

Data Paper

# Building a reliable 16S mini-barcode library of wild bees from Occitania, south-west of France

Anaïs Marquisseau[‡], Kamila Canale-Tabet[§], Emmanuelle Labarthe[§], Géraldine Pascal[§], Christophe Klopp[|], André Pornon[¶], Nathalie Escaravage[¶], Rémi Rudelle[#], Alain Vignal[§], Annie Ouin[‡], Mélodie Ollivier[‡], Magalie Pichon[‡]

‡ Dynafor, INRAE, INP, ENSAT, 31326, Castanet Tolosan, France
§ GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet Tolosan, France
| MIAT, INRAE, 31326, Castanet Tolosan, France
¶ CRBE, CNRS, UPS, IRD, INP, 31077, Toulouse, France
# Rudelide Expertise muséologie, Rieupeyroux, France

Corresponding author: Anaïs Marquisseau (anais.marquisseau@inrae.fr), Magalie Pichon (magalie.pichon@inrae.fr)

Academic editor: Benoît Geslin

Received: 20 Sep 2024 | Accepted: 12 Dec 2024 | Published: 07 Jan 2025

Citation: Marquisseau A, Canale-Tabet K, Labarthe E, Pascal G, Klopp C, Pornon A, Escaravage N, Rudelle R, Vignal A, Ouin A, Ollivier M, Pichon M (2025) Building a reliable 16S mini-barcode library of wild bees from Occitania, south-west of France. Biodiversity Data Journal 13: e137540.
https://doi.org/10.3897/BDJ.12.e137540

## Abstract

### Background

DNA barcoding and metabarcoding are now powerful tools for studying biodiversity and especially the accurate identification of large sample collections belonging to diverse taxonomic groups. Their success depends largely on the taxonomic resolution of the DNA sequences used as barcodes and on the reliability of the reference databases. For wild bees, the barcode sequences coverage is consistently growing in volume, but some incorrect species annotations need to be cared for. The COI (Cytochrome Oxydase subunit 1) gene, the most used in barcoding/metabarcoding of arthropods, suffers from primer bias and difficulties for covering all wild bee species using the classical Folmer primers.

© Marquisseau A et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## New information

We present here a curated database for a 250 bp mini-barcode region of the 16S rRNA gene, suitable for low-cost metabarcoding wild bees in applications, such as eDNA analysis or for sequencing ancient or degraded DNA. Sequenced specimens were captured in Occitania (south-west of France) and morphologically identified by entomologists, with a total of 530 individuals belonging to 171 species and 19 genera. A customised workflow including distance-tree inferences and a second round of entomologist observations, when necessary, was used for the validation of 348 mini-barcodes covering 148 species. Amongst them, 93 species did not have any 16S reference barcode available before our contribution. This high-quality reference library data are freely available to the scientific community, with the aim of facilitating future large-scale characterisation of wild bee communities in a context of pollinators' decline.

## Keywords

wild bees, *Apoidea*, Anthophila, 16S rRNA, reference database, DNA barcoding

## Introduction

Worldwide, pollinators have become the focus of particular attention as populations decline drastically (Biesmeijer et al. 2006, Rhodes 2018, Powney et al. 2019). Amongst these, wild bees (Hymenoptera, *Apoidea*, Anthophila) provide the majority of pollination services with more than 20,000 species listed on the Planet (Michener 2007). Many countries have launched important research programmes in order to define actions required for their conservation and restoration, such as ORBIT or RestPoll in Europe. In France, the establishment of the IUCN Red List of wild bees as a part of the current Pollinator Plan (2021-2026) is a major action to counteract the decline of pollinators. In this context, extensive temporal and spatial sampling is crucial for ecological and conservation studies and needs to be associated with rapid and cost-effective identification of large specimens numbers.

Traditionally, arthropod identification, including wild bees, was based on the examination of morphological characters and the time-consuming detection of subtle morphological differences between species requires trained taxonomists. Unfortunately, the lack of policy commitment to training new experts has led to an increasingly intense shortage of specialists, a situation commonly referred to as the taxonomic impediment (de Carvalho et al. 2007, Vinarski 2020). To complement these morphology-based methods, molecular approaches have been developed, rapidly becoming essential tools in modern taxonomy (Chua et al. 2023). In 2007, the Barcode of Life project was initiated in Ontario with the objective of creating a public reference library for all animal species of a standard 640 bp fragment from the mitochondrial COI (Cytochrome Oxydase subunit 1) gene (Ratnasingham and Hebert 2007, Ratnasingham and Hebert 2013). Since then, the COI

marker has been widely used for DNA metabarcoding purposes, particularly to describe arthropod biodiversity in various contexts (Piper et al. 2019, Liu et al. 2019, Remmel et al. 2024).

As a consequence, the COI barcode has become the main marker used for cataloguing the genetic diversity of *Apoidea* Anthophila in many countries worldwide: in Canada (Sheffield et al. 2009, Sheffield et al. 2017), in Chile (Packer and Ruz 2017), in Ireland (Magnacca and Brown 2012), in United Kingdom (Creedy et al. 2020), in Germany (Schmidt et al. 2015), in Luxembourg (Herrera-Mesías et al. 2022), in Slovenia (Janko et al. 2024), in Spain and Portugal (Wood et al. 2024) and in France (Villalta et al. 2021, Ollivier et al. In prep.). However, some of these studies have identified difficulties in efficiently barcoding some wild bee species such as *Andrena* or *Hylaeus* with the classical COI Folmer primers (658 bp) (Folmer et al. 1994, Schmidt et al. 2015, Villalta et al. 2021). An additional difficulty arises from the existence of a few inaccurate wild bees species annotations in the BOLD (Ratnasingham et al. 2024) and GenBank (Clark et al. 2016) databases, such as reported by Herrera-Mesías et al. (2022) and Janko et al. (2024). Thus, for the success of future metabarcoding investigations, there is a need for the evaluation of other barcode sequences and their potential to ensure the largest possible taxonomic coverage of wild bees species and to improve database curation. Amongst the possible candidate mitochondrial genes usable for low-cost wild bee metabarcoding, the 16S rRNA gene could be a good choice because of its short highly variable and conserved regions (Marquina et al. 2018, Elbrecht et al. 2016).

Since over two decades, the 16S locus has already been used to infer the phylogeny of Hymenoptera including bees (Whitfield and Cameron 1998). Molecular phylogenies of Apoidae with 16S rRNA were reported for stingless bees living in Neotropical regions (Costa et al. 2003, Trianto and Purwanto 2020, Marconi et al. 2023) and for honey bees subspecies in Saudi Arabia (Alajmi et al. 2019). With primers derived from the *Apis mellifera* 16S sequences, the phylogeny of some Korean bumblebees was clarified (Yoon et al. 2004). Kek et al. (2017) tested two short regions of both the COI and 16S genes to discriminate bee species involved in honey production and demonstrated that a 287 bp region of the 16S rRNA gene was more informative than a 284 bp region of the COI gene. Moreover, a 16S mini-barcode (120 bp long) has been tested with success on Insecta class by Hsieh et al. (2019).

Targeting a short barcode gene region (hereafter referred to as mini-barcode) is particularly interesting for approaches requiring to overcome DNA degradation, while preserving a high level of taxonomic resolution (Hajibabaei et al. 2006). Amongst these approaches, sequencing museum specimens (Levesque-Beaudin et al. 2023, Santos et al. 2023) or eDNA based biomonitoring provide encouraging prospects (Newton et al. 2023, Sickel et al. 2023, Avalos et al. 2024). Currently, there is no mini-barcode library available for wild bee species, unlike the ones already available for plants (Little 2014) and marine macrophytes (Ortega et al. 2020).

In this study, our main objectives were to evaluate the 16S mini-barcode potential (Clarke et al. 2014) to discriminate wild bee species and build a robust database facilitating future DNA metabarcoding investigations on these important pollinators.

## Sampling methods

### Step description: Collection description

The 530 specimens used in this study originated from three sources: 1) 412 from the UMR DYNAFOR collection stored at INP-ENSAT (Ollivier et al. 2024); 2) 88 from the private collection of the bee expert, Rémi Rudelle and 3) a set of 30 *Bombus* specimens from the CRBE (Centre de Recherche sur la Biodiversité et l'Environnement) collection. Metadata with detailed information related to each specimen (geographic location, identifiers, sex etc.) can be found in Suppl. material 1.

### 1. Sample collection

For the UMR DYNAFOR collection, three coloured pan traps (blue, white and yellow) were set in the grassy strip boarding the crop. Each pan trap contained 250 ml of water with Teepol (3 drops/l). After 3 days of exposure, the insects were filtered and placed in ethanol (EtOH) 70% until identification. A panel of 61 specimens was captured with nets and euthanised in ethyl acetate vapour (Suppl. material 1). For the private collection of Rémi Rudelle and the CRBE collection, the specimens were captured with nets.

All of the specimens were morphologically identified using mostly the Insecta Fauna Helvetica reference (Amiet 1996, Amiet 1999, Amiet et al. 2001, Amiet et al. 2004, Amiet et al. 2007, Amiet et al. 2010, Amiet et al. 2017) and others (Schmid-Egger and Scheuchl (1997), Wood (2023) for the Andrenidae family; Ortiz-Sanchez and Jimenez-Rodriguez (1991), Terzo et al. (2007), Rasmont (2014), Smit (2018), Aubert (2020), Le Divelec (2021), Rasmont et al. (2021) for the Apidae family; Ornosa and Ortiz-Sanchez (2004) for the Colletidae family; Pauly (2019) for the Halictidae family and Benoist (1931), Benoist (1941), Pauly (2015) for the Megachilidae family) by one of the following entomologists: Rémi Rudelle, David Genoud, Romain Carrié, Léa Frontero and Dominique Pelletier. They are conserved, pinned in insect boxes and stored at room temperature. To prevent parasite infestation, specimens are frozen twice a year at -20°C for at least 48 hours for the UMR DYNAFOR collection.

### 2. Sequencing and processing

DNA was extracted from a portion or an entire leg of dried specimens using the Chelex method (see Casquet et al. (2012) for a detailed protocol). Two sequencing technologies were used: 275 specimens were sequenced using high-throughput Illumina technology (MiSeq Sequencing System) and 255 were sequenced with Sanger technology. The list of all species with the corresponding sequencing method is included in Suppl. material 1.

MiSeq sequencing and processing

For the set of 275 specimens processed with MiSeq sequencing, two microlitres of DNA were used as template for PCR. 16S primers ins16S_1R/ins16S_1F (R: TRRGACGAGAAGACCCTATA; F: TCTTAATCCAACATCGAGGTC, Clarke et al. (2014)) were chosen to amplify a 250 bp region of the mitochondrial 16S gene. PCR was performed in a 20 µl total volume containing 5.84 µl of purified water, 10 µl of 2x ampliTaq Gold 360 master mix (Thermo Scientific LSG Life Technologies) including dNTP and ampliTaq Gold, 0.16 µl BSA, 1 µl of each primer (5 µM) and 2 µl of DNA. PCR was carried out under the following conditions: hot-start at 95°C for 10 min followed by 40 cycles (denaturation at 95°C for 30 s, primer annealing at 49°C for 30 s and primer extension at 72°C for 30 s); and final extension at 72°C for 7 min. Primers were 5′ labelled with a set of 8 bp tags to identify samples in bioinformatics analysis. 16S PCR products were visualised on 1% TAE agarose gels quantified using PicoGreen dsDNA Quantitation Reagent and mixed aiming at equimolar pools. The pool was then purified using beads contained in the Illumina TruSeq Nano kit (part #15041758) and libraries were generated following the manufacturer's guide for the Illumina TruSeq Nano kit, except that no sonication was performed. Libraries were sequenced on a single run of an Illumina MiSeq (2 × 250 paired-end reads), using the NGS core facility at the Génopole Toulouse Midi-Pyrénées. We obtained 22,540,200 demultiplexed reads (R1 and R2 reads). 16S rDNA amplicon sequences were analysed using the FROGS pipeline (version 3.1, Escudié et al. (2017)). Amplicons were processed according to their size (150 - 490 nucleotides) and clustered into ASVs (Amplicon Sequence Variant) using Swarm (aggregation distance: d = 1) (Mahé et al. 2014). For each sample, the most abundant ASV was kept for the procedure of barcode validation.

Sanger sequencing and processing

DNA barcoding using Sanger sequencing technology was performed on 255 specimens. Specific primers were used for each genus. All primer sequences and PCR conditions are given in Fig. 1. For each PCR reaction, 3 µl of extracted DNA was amplified in 25 µl final volume, 1 µM for each primer, 1 x PCRBIO Reaction Buffer (including Mg and dNTPs) and 0.25 µl of PCRBIO Taq DNA Polymerase (5 u/µl) (PB10.11-20; Eurobio). Prior to sequencing, a volume of 2.5 µl from each PCR product was examined on a 2% agarose gel electrophoresis to check the success and specificity of the PCR amplification. The sequencing reaction was subsequently prepared as follows: 2.5 µl of each PCR product was purified by adding 1 µl of each Exonuclease (M0293L; Ozyme) and TSAP (Thermo Sensitive Alcaline Phosphatase) (M9910; Promega) in a final volume of 18 µl. The sequencing reaction mixture was split in two volumes and 1 µl of 10 mM of each primer (forward and reverse) was added. PCR products were sent to a private company for Sanger sequencing in both directions. The sequences produced were manually checked for base calling using ChromasPro 2.1.10.1. (Technelysium Pty Ltd, Tewantin, Australia) and unreadable sequences were removed. For the 30 *Bombus* sequences from CRBE laboratory, the PCR was performed with the forward primer: CGCTGTTATCCCTAAGG and the reverse primer CTGTACAAAGGTAGCATAATC.

| Genus | Primer Set | Forward Primer (5'-3') | Reverse Primer (5'-3') | No. of cycles | Cyclic denaturation, annealing and elongation (temp./duration) |
|---|---|---|---|---|---|
| *Andrena* | A | TTAATTCAACATCGAGGTCGC | GGGACGAGAAGACCC | 38 | 95°C/45sec<br>60°C/45sec<br>72°C/45sec |
| *Bombus* | C | CTTAATTCAACATCGAGGTC | GGGACGATAAGACCC | 35 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |
| *Ceratina* | E | TTAATTCAACATCGAGGTCGC | GGGACGATAAGACCC | 35 | 95°C/45sec<br>50°C/45sec<br>72°C/45sec |
| *Chelostoma* | D | TTAATTCAACATCGAGGTCGC | GGGACGATAAGACCC | 35 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |
| *Eucera* | E | TTAATTCAACATCGAGGTCGC | GGGACGATAAGACCC | 35 | 95°C/45sec<br>50°C/45sec<br>72°C/45sec |
| *Halictus* | C | CTTAATTCAACATCGAGGTC | GGGACGATAAGACCC | 38 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |
| *Hoplitis* | D | TTAATTCAACATCGAGGTCGC | GGGACGATAAGACCC | 35 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |
| *Hylaeus* | A | TTAATTCAACATCGAGGTCGC | GGGACGAGAAGACCC | 35 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |
| *Lasioglossum* | B | CTTAATTCAACATCGAGGTC | TRRGACGAGAAGACCCTATA | 38 | 95°C/45sec<br>50°C/45sec<br>72°C/45sec |
| *Nomada* | D | TTAATTCAACATCGAGGTCGC | GGGACGATAAGACCC | 38 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |
| *Osmia* | D | TTAATTCAACATCGAGGTCGC | GGGACGATAAGACCC | 35 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |
| *Seladonia* | C | CTTAATTCAACATCGAGGTC | GGGACGATAAGACCC | 38 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |
| *Sphecodes* | D | TTAATTCAACATCGAGGTCGC | GGGACGATAAGACCC | 35 | 95°C/45sec<br>58°C/45sec<br>72°C/45sec |

Figure 1. doi

**Primer sequences and PCR conditions used for Sanger sequencing per Genus.** For all Sanger samples, primary denaturation was performed at 95°C for 5 minutes and final elongation was performed at 72°C for 20 minutes.

Amongst the 171 species included in our study, 43 were represented by a single specimen, 25 by two specimens, 51 by three specimens and 52 by four to ten specimens, corresponding to a total of 530 specimens (Fig. 2, step 1).

The global success rate after MiSeq sequencing was very high, reaching 99%, with only a single specimen failing. The sequencing success with the Sanger technology was lower due to negative PCR or unreadable chromatograms. Indeed, no sequence could be obtained for 66 specimens. However, as replicate samples were included for most species, only seven species were excluded (Suppl. material 3) at this step (Fig. 2, step 2).

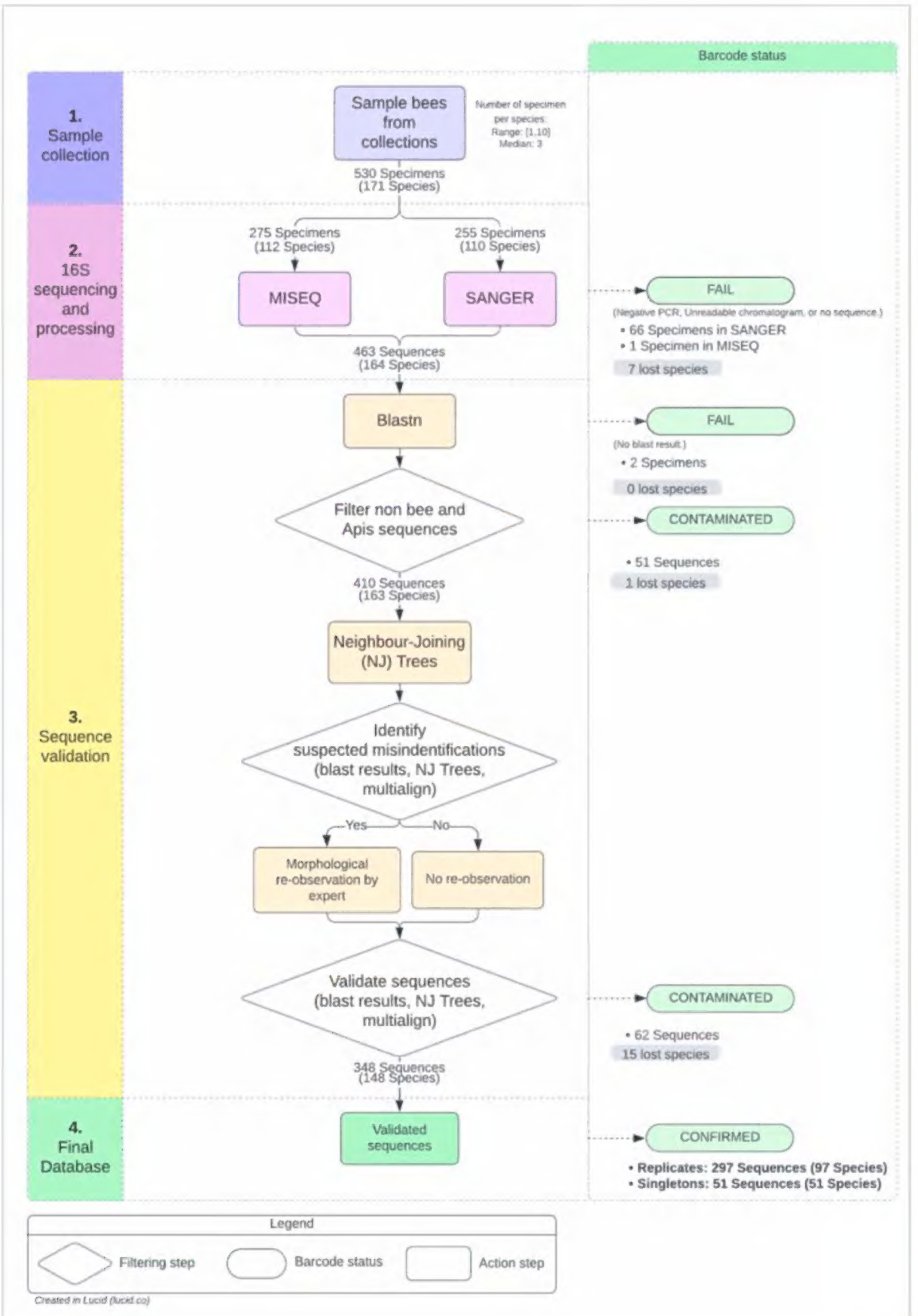


Figure 2. doi

16S mini-barcode library workflow, from sampling to validation. There are four main steps represented by different colours in the left panel, from top to bottom. The middle part of the figure represent the workflow. The right panel indicates the final barcode status.

For sequence validation, we used : sequence assignation by BLASTn (Altschul et al. 1990) on GenBank nt database (Sayers et al. 2021); Neighbour-joining (NJ) distance-tree inferences using the K2P model and the Muscle algorithm (Edgar 2004) for alignment implemented in the BOLD toolkit; Sequence alignment using multalin software to visualise allelic variations (Corpet 1988). For some species, the default BLAST parameters were adjusted to take into account for the high AT content of the region in this genus.

### 3. Sequence validation

The 463 sequences corresponding to the remaining 164 species were searched in GenBank (nt database) using BLASTn (Altschul et al. 1990) and contaminants (sequences which do not match with a wild bee reference) were removed from data. Fifty-three non-bee and *Apis* sequences were eliminated and the 410 remaining sequences (163 species) were analysed by cross validation with two filtering rounds (Fig. 2, step 3). The first round allows the detection of potential misidentifications as incongruence between the morphologically and the genetically-based species identification through Neighbour-Joining Trees. For species with only one specimen (singleton), the barcode was validated if the sequence belonged to the corresponding genus. Potentially misidentified species, as well as species that are known to be part of a species complex, were submitted to an entomologist for a second observation taking into account possible identification key updates. The second filtering round allowed sequences validation and confirmation as accurate barcodes. As a result, 16 additional species were excluded (Suppl. material 3).

In total 348 sequenced samples corresponding to 148 unique species were successfully analysed and validated and 97 species were represented by at least two specimens (Fig. 2, step 4).

## Geographic coverage

**Description:** The wild bees presented in this study were collected from the French region of Occitanie (Fig. 3). The 412 specimens coming from the UMR DYNAFOR collection were collected in 17 sites located in south-west of France, in the long term socio-ecological research site Zone Atelier Pyrénées-Garonne (ZA PYGAR, Ouin et al. (2021)) over a period of 7 years (2013-2019). The ZA PYGAR takes place in the Pyrénées foothills and is characterised by a mosaic of landscapes with crops and small forests (Carrié et al. 2018, Rivers-Moore et al. 2020). Eighty-eight specimens were captured by Rémi Rudelle in different sites of the Aveyron, French Department and 30 *Bombus* specimens from the CRBE collection were sampled in the Pyrénées Orientales, French Department (personal communication Nathalie Escaravage).

## Taxonomic coverage

**Description:** Specimen records are reported for the 348 sequences (148 species) confirmed with the above workflow. Fig. 4 displays the sequences and species covered by the herein presented 16S library for each genus.

For 55 of the species included in our dataset, 16S partial or full-length sequences were already available in GenBank (Fig. 4). The list and the accession number of these is reported in Suppl. material 4. The *Bombus* genus is the most represented with 22 species and 118 sequences. For *Andrena*, 26 16S sequences corresponding to 15 species

originated from the mitochondrion sequencing project were available. Seven species (22 sequences) of *Lasioglossum* were available at the time of writing the manuscript. At the end, there were no 16S data for 10 genera of wild bees. Thus, we provide 204 new 16S mini-barcodes for wild bees belonging to 93 species. For the most abundant species of France belonging to the *Andrena* and *Lasioglossum* genera sets, 71 new sequences (32 species) and 37 new sequences (18 species) were respectively added in the public databases.

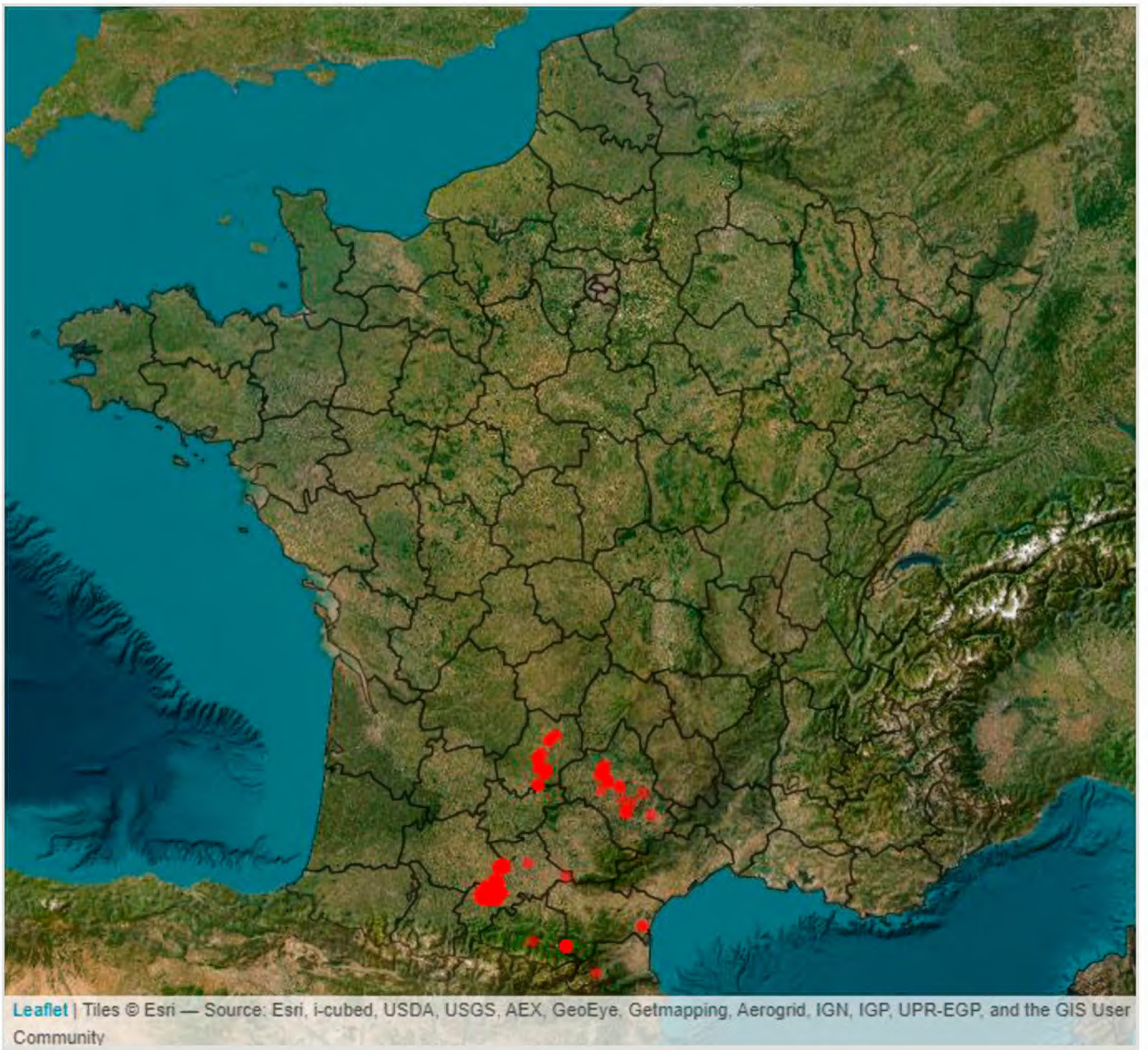
Leaflet | Tiles © Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community

Figure 3. doi

Geographic distribution of the 16S database specimens collected in Occitanie (red points). Nine specimens are not mapped as they were collected outside of Occitanie, although they represent species that can be found within the region.

## Temporal coverage

**Data range:** 2010-5-18 - 2020-7-22.

| Genus per Family | 16S Sequence from local database nb seq (nb sp) | 16S Sequence from Genbank nb seq (nb sp) | New 16S barcode available nb seq (nb sp) |
|---|---|---|---|
| ***Andrenidae*** | | | |
| *Andrena* | 111 (47) | 26 (15) | **71 (32)** |
| ***Apidae*** | | | |
| *Anthophora* | 3 (1) | 1 (1) | 0 (0) |
| *Bombus* | 73 (23) | 118 (22) | **5 (1)** |
| *Ceratina* | 6 (2) | 0 (0) | **6 (2)** |
| *Eucera* | 17 (6) | 0 (0) | **17 (6)** |
| *Nomada* | 21 (13) | 10 (6) | **13 (7)** |
| *Tetralonia* | 3 (1) | 0 (0) | **3 (1)** |
| *Xylocopa* | 5 (3) | 0 (0) | **5 (3)** |
| ***Colletidae*** | | | |
| *Colletes* | 1 (1) | 0 (0) | **1 (1)** |
| *Hylaeus* | 8 (5) | 0 (0) | **8 (5)** |
| ***Halictidae*** | | | |
| *Halictus* | 23 (8) | 0 (0) | **23 (8)** |
| *Lasioglossum* | 51 (25) | 22 (7) | **37 (18)** |
| *Seladonia* | 6 (3) | 2 (1) | **3 (2)** |
| *Sphecodes* | 1 (1) | 0 (0) | **1 (1)** |
| ***Megachilidae*** | | | |
| *Chelostoma* | 1 (1) | 0 (0) | **1 (1)** |
| *Hoplitis* | 1 (1) | 0 (0) | **1 (1)** |
| *Osmia* | 17 (7) | 8 (3) | **9 (4)** |

Figure 4. doi

Taxonomic coverage of the 16S mini-barcode library compared to existing sequences in GenBank. nb seq: Number of sequences ; nb spe: Number of species. As a point of reference, the sequences available on the public database GenBank before the project start are given in the second column (only for species recorded in the area of collection ZA PYGAR). The last column indicates the number of new sequences added in GenBank and BOLD. Number of species corresponding to specimen records are indicated in brackets.

## Usage licence

**Usage licence:** Other

**IP rights notes:** CC BY 4.0

## Data resources

**Data package title:** 16S mini-barcode library of wild bees from Occitania

**Resource link:** dx.doi.org/10.5883/DS-BCWBS16S

**Number of data sets:** 1

**Data set name:** DS-BCWBS16S

**Download URL:** https://v4.boldsystems.org/index.php/Public_BINSearch?searchtype=records

**Data format:** tsv, fasta

**Description:** The list of the 530 specimens (171 species) with complementary information such as their BOLD IDs, process IDs, GenBank IDs (only for sequences > 200 bp), taxonomy, identifiers, gps location for UMR DYNAFOR collection, sequencing method, barcode status (failed, contaminated or confirmed replicate/ single) is contained in the dataset. It covers five families, 19 genera and 171 species. After sequencing and validation barcode steps, 348 sequences corresponding to 148 species and 17 genera were selected. Suppl. materials 1, 2 can be downloaded as the version of the dataset (metadata and fasta sequences) at the time of writing the manuscript.

| Column label | Column description |
|---|---|
| Sample_ID | Unique BOLD identifier for the specimen. |
| Process_ID | Unique BOLD identifier for the barcode. |
| Accession_NCBI | Unique GenBank identifier for the barcode (Accession number). |
| Museum_ID | Unique collection identifier for the specimen. |
| Collection_code | Identifier for the collection: Dynafor, RIEUPEYROUX or CRBE. |
| Institution_storing | Institution where specimens are physically stored: ENSAT, Rudélide Expertise Muséologie REM or CNRS. |
| Phylum | Phylum name |
| Class | Class name |
| Order | Order name |
| Family | Family name |
| Subfamily | Subfamily name |
| Genus | Genus name |
| Species | Species name |
| Subspecies | Subspecies name |
| Identifier | Name of the individual who identified the specimen morphologically. |
| Identifier_Email | Email of the identifier. |
| Identification_Method | All specimens were morphologically identified. |
| Sex | The sex of the specimen: F for female, M for male. |

| | |
|---|---|
| Specimen's caste | Extra information about the specimen's caste: W if the specimen is a worker (empty otherwise). |
| Life_stage | Life stage of the sampled specimen. All specimens were adults. |
| Tissue_descriptor | Type of tissue analysed: LEG. |
| Collectors | Names of the individuals who collected the specimen in the field. |
| Collection_Date | Exact date during which the specimen was collected. For CRBE specimens, only the collection year is available. |
| Country | Name of the country in which the specimen was collected. All specimens were collected in France. |
| State | Name of the state (French: Région) in which the specimen was collected. All the specimens, except nine, were collected in Occitanie. |
| Region | Name of the region (French: Département) in which the specimen was collected. |
| Sector | Name of the sector or city, in which the specimen was collected. |
| Exact_Site | A brief description of the site in which the specimen was collected. |
| Latitude | The geographic latitude (in decimal degrees) of the site in which the specimen was collected. For CRBE or Rudélide specimens, only an approximate latitude is available, corresponding to the latitude of the municipality rather than the exact collection point. |
| Longitude | The geographic longitude (in decimal degrees) of the site in which the specimen was collected. For CRBE or Rudélide specimens, only an approximate longitude is available, corresponding to the longitude of the municipality rather than the exact collection point. |
| Sampling_protocol | The sampling method used to capture the specimen: NET or PAN TRAP. |
| Sequencing_method | The method used to sequence the specimen: SANGER or MISEQ. |
| Barcode_status | The status of the 16S barcode for the specimen: Confirmed_single, Confirmed_replicate, Contaminated or Fail. |

# Additional information

## Sequencing and barcoding results

The within-genus global sequencing success including MiSeq and Sanger technologies varies from 75% to 100%, except for *Sphecodes* (33%) (Fig. 5). For the most represented genus of our dataset, namely *Andrena* and *Lasioglossum*, the sequencing success rates were 84% and 78%, respectively. The barcoding success rate after all successive filtering steps was different according to genera. It was 100% for *Nomada*, *Tetralonia*, *Antophora* and *Colletes*, 90% for *Bombus*, 88% for *Hylaeus*, 77% for *Osmia*, 76% for *Halictus*, 72% for *Andrena* and it was under 65% for the other genera. For *Lasioglossum*, the low barcoding success rate (38%) is related to low sequencing success, being as low as 42.7% with MiSeq and 28% with Sanger (Suppl. material 1) and to high level of

contaminated sequences (54). An in-depth analysis of *Lasioglossum* sequences showed that the end of the amplified fragment contains many stretches of AT nucleotide repeats (Suppl. material 2) which are known to disrupt the polymerase activity during the sequencing process. Amongst the 171 species of our dataset, no barcode could be obtained for 23 species including nine species of *Lasioglossum* (Suppl. material 3). Suppl. material 5 provides detailed sequencing and barcoding success per species.

| Genus per Family | Number of sampled specimens | Number of confirmed barcodes | Number of contaminated sequences | Number of specimens without a sequence (Fail) | Sequencing Success (%) | Barcoding Success (%) |
|---|---|---|---|---|---|---|
| **Andrenidae** | | | | | | |
| *Andrena* | 154 | 111 | 19 | 24 | 84 | 72 |
| **Apidae** | | | | | | |
| *Anthophora* | 3 | 3 | 0 | 0 | 100 | 100 |
| *Bombus* | 81 | 73 | 6 | 2 | 97 | 90 |
| *Ceratina* | 10 | 6 | 3 | 1 | 90 | 60 |
| *Eucera* | 26 | 17 | 5 | 4 | 84 | 65 |
| *Melecta* | 1 | 0 | 1 | 0 | 100 | 0 |
| *Nomada* | 21 | 21 | 0 | 0 | 100 | 100 |
| *Tetralonia* | 3 | 3 | 0 | 0 | 100 | 100 |
| *Xylocopa* | 8 | 5 | 3 | 0 | 100 | 62 |
| **Colletidae** | | | | | | |
| *Colletes* | 1 | 1 | 0 | 0 | 100 | 100 |
| *Hylaeus* | 9 | 8 | 0 | 1 | 88 | 88 |
| **Halictidae** | | | | | | |
| *Halictus* | 30 | 23 | 6 | 1 | 96 | 76 |
| *Lasioglossum* | 135 | 51 | 55 | 29 | 78 | 37 |
| *Seladonia* | 11 | 6 | 4 | 1 | 90 | 54 |
| *Sphecodes* | 6 | 1 | 1 | 4 | 33 | 16 |
| **Megachilidae** | | | | | | |
| *Anthidium* | 1 | 0 | 1 | 0 | 100 | 0 |
| *Chelostoma* | 4 | 1 | 3 | 0 | 100 | 25 |
| *Hoplitis* | 4 | 1 | 2 | 1 | 75 | 25 |
| *Osmia* | 22 | 17 | 4 | 1 | 95 | 77 |

Figure 5. doi

Sequencing and barcoding success per family and genus.

## Analysis of genetic distances

Examination of the general normalised divergence histogram performed with BOLD analyses toolkit on all species (replicates) indicates the existence of a DNA barcoding gap (maximal intraspecific distance > minimal interspecific distance), allowing reliable molecular identification of specimens (Fig. 6). However, a more in-depth analysis of each genus reveals two scenarios: There was an overlap between intraspecific and interspecific genetic distances in six genera: *Andrena*, *Bombus*, *Eucera*, *Halictus*, *Lasioglossum* and *Nomada*, whereas the barcoding gap was clearly existing for the five others genera: *Xylocopa*, *Seladonia*, *Hylaeus*, *Osmia* and *Ceratina* (Fig. 6). The tables with detailed intraspecific and interspecific genetic distances are given in Suppl. materials 6, 7.

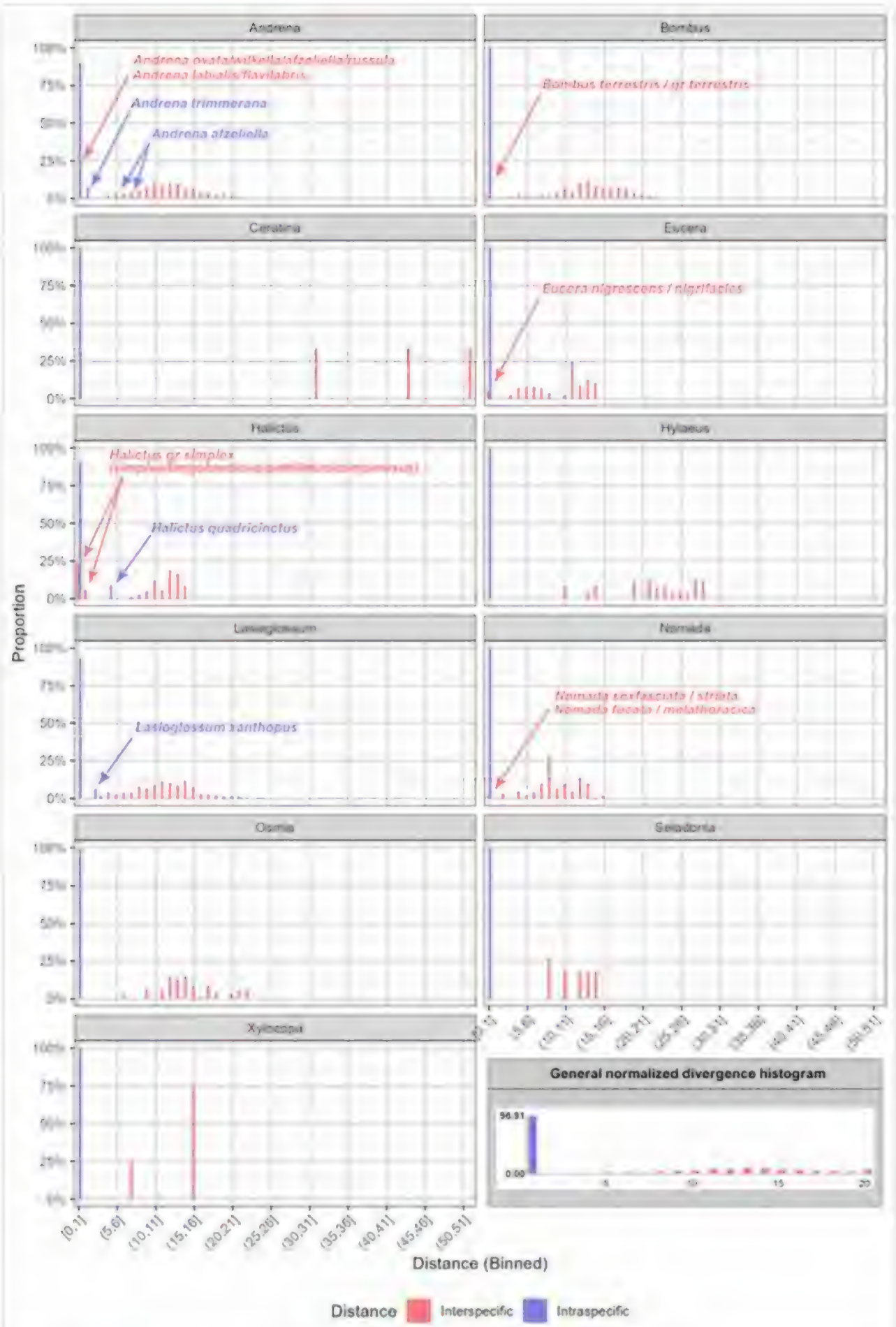


Figure 6. doi

Distribution of intraspecific and interspecific genetic distances per genus. The global normalised distance distribution for all specimens is shown at bottom right corner. Blue arrows indicate species that show an intraspecific distance > 1%. Red arrows indicate group of species that show an interspecific distance < 1%.

**Genetic distance analyses per family and genus**

- **Andrenidae**: (Taxon ID tree is given in Suppl. material 8).

For the *Andrena* genus, which has been reported difficult to barcode with the COI Folmer primers (Schmidt et al. 2015, Villalta et al. 2021), the 16S mini-barcode offers a good alternative. Amongst the 111 validated barcodes (47 species), the 16S mini-barcode allows us to discriminate all the species in accordance with the morphological subgenera classification (Suppl. material 8). Interestingly, the 250 bp of the 16S gene used in this study is sufficiently divergent to separate complex groups previously described in literature. For example, the barcoding of *Andrena distinguenda* species group (*A. nitidula* and *A. distinguenda*) with COI showed the existence of two bins (Schmidt et al. 2015). With the 16S mini-barcode, the minimum divergence between these two species was 1.13% supporting the existence of two species (Suppl. material 7). Similar to that described by Wood et al. (2021), we found a clear separation (3.11% minimum divergence) in the *Andrena angustior* group between *Andrena impressa* and *Andrena fulvata* (Suppl. material 7). A complex situation remains with *Andrena trimmerana*; our molecular data on eight specimens including two males and six females clearly show two groups (0% divergence within groups) with an intergroup divergence of 1.29% (Suppl. material 7). Interestingly, we observed allelic variation (1 SNP) between *Andrena dorsata* originating from our data and the two *Andrena dorsata* sequences provided in GenBank originated from the UK (KT16433.1 and OV815490.1). Elsewhere, two sequences of *Andrena fulva* were available in GenBank (KT164623.1 and OX276334.1). Alignment of these two *A. fulva* with our specimens reviewed by entomologists show that KT16423.1 is 100% homologous with our *Andrena fulvago*, whereas OX276334.1 aligns with *Andrena fulva*. The sequences of these two species diverge 5.9% with the 16S mini-barcode.

- **Apidae**: (Taxon ID tree is given in Suppl. material 9).

***Bombus***: The distance tree inferred from the 16S mini-barcodes of the 23 *Bombus* species (73 specimens) reveals genetic divergence that is consistent with the known subgenus classification (Cameron et al. 2007, Cejas et al. 2019, Sun et al. 2021). Interestingly, the 16S mini-barcode classifies without ambiguity each specimen of the following species complex: *Bombus pascuorum*/*muscorum*; *Bombus sylvarum*/*ruderarius*; *Bombus ruderatus*/*hortorum*; *Bombus pyreaneus*/*pratorum* and *Bombus terrestris*/ *lucorum*. As some of the specimens in the *Bombus terrestris* group in our dataset were not morphologically identified at the species level, the 0% divergence between specimens named *Bombus* gr. *terrestris* and *Bombus terrestris* suggests that all of them are *Bombus terrestris*. Elsewhere, amongst the 10 *Bombus pascuorum* specimens barcoded with the 16S mini-barcode, some allelic variation was observed, with a minimum intraspecific divergence of 0% and a maximum intraspecific divergence of 0.98% (Suppl. material 6). In Switzerland, Amiet et al. (2017) reported the presence of two subspecies *of Bombus pascuorum*.

***Eucera***: In the Apidae family, the 16S mini-barcodes are not discriminant for two species belonging to the *Eucera* genus: *Eucera nigrifacies* and *Eucera nigrescens*, whereas it

efficiently delineates the six others species, especially *Eucera longicornis* which was confused in the past with *Eucera nigrescens* (Dorchin et al. 2018). Allelic variations are observed for *Eucera longicornis* (0.26% maximum intraspecific divergence) and *Eucera numida* (0.26% maximum intraspecific divergence) (Suppl. material 6).

***Nomada***: Regarding the *Nomada* genus, recently reexamined by Odanaka et al. (2022) and Straka et al. (2024), our data shows that the 16S mini-barcode distinguishes the 21 specimens (13 species), except *Nomada striata* versus *Nomada sexfasciata* and *Nomada fucata* versus *Nomada melathoracica*. However, more specimens need to be analysed to conclude definitively that three of these species are singletons.

***Ceratina***: The molecular phylogeny of these small carpenter bees has been recently achieved by Sless et al. (2024). The two species of our dataset are extremely divergent (31.12% min interspecific divergence and 51.49% max interspecific divergence, Suppl. material 7). One belongs to the subgenus *Euceratina* (*Eucera cyanea*) and the other to the subgenus *Ceratina sensu stricto* (*Ceratina cucurbitinia*).

***Xylocopa***: The five specimens of *Xylocopa* from our dataset correspond to the three species: *Xylocopa valga*, *Xylocopa iris* and *Xylocopa violacea* exhibit 0.46% maximum intraspecific divergence and 7.19% minimum interspecific divergence (Suppl. materials 6 , 7).

- **Colletidae** (Taxon ID tree is given in Suppl. material 10).

In the present study, five species belonging to *Hylaeus* (Almeida and Danforth 2009) genus were successfully barcoded with the 16S mini-barcode. Amongst them, four were singletons and *Hylaeus brevicornis* had three replicates with intraspecific divergence of 0% (Suppl. material 6). The divergence between species ranges from 10.47% to 28.66% (Suppl. material 7).

- **Megachilidae** (Taxon ID tree is given in Suppl. material 11).

***Osmia***: As *Apis mellifera* or *Bombus*, *Osmia* are commercially reared for pollination services. A complete phylogeny of the Palearctic Osmiine bee is available on the website of Müller (2024). Molecular data using UCEs or Elongation factor 1-α or LW-rhodopsin and Conserved ATPase domain have been reported by Praz et al. (2008) and Branstetter et al. (2021). In this work, the 16S mini-barcode is clearly efficient to separate the seven species of our dataset: 0.51% maximum intraspecific divergence and 6.22% minimum interspecific divergence were observed (Suppl. materials 6, 7).

- **Halictidae** (Taxon ID tree is given in Suppl. material 12).

***Lasioglossum***: Molecular phylogeny of *Lasioglossum* is poorly documented (Danforth 1999, Gibbs et al. 2012, Gibbs 2018, Pauly et al. 2019). For the 25 species belonging to genus *Lasioglossum* successfully barcoded in this study, the intraspecific divergence is < 1% for all species, except for *Lasioglossum xanthopus* which is > 2%. Interestingly, the 16S mini-barcode allows a clear separation (min interspecific divergence > 2%) for

complex groups. Thus, for the specimens of *Lasioglossum medinail/Lasioglossum villosulum* species, the min interspecific divergence is 4.38%. It is 6.84% for *Lasioglossum malachurum/subhirtum/calceatum/pauxillum/laticeps* species; 5.70% for *Lasioglossum morio/nitidulum* species and 5.23% for *Lasioglossum pauperatum/ pygmaeum/truncaticolle/crassepunctatum* species (Suppl. materials 6, 7).

***Halictus***: In the *Halictus simplex* group, *Halictus simplex* and *Halictus langobardicus* females are extremely difficult to distinguish morphologically. Unfortunately, the 16S mini-barcode did not allow for the discrimination of the two species, whereas *Halictus compressus* exhibited a 0.78% minimum interspecific divergence with the rest of the group (Suppl. material 7). Interestingly, two specimens from the *Halictus simplex* group diverged slightly from the others. It would be interesting to barcode more specimens including species belonging to the *simplex* group which were not included in our dataset. Sequencing complete mitochondrion in the future would help to clarify the status of this group. Elsewhere, we observed allelic variation amongst specimens of *Halictus quadricinctus* and we suspected that one of them could be *Halictus brunnescens*.

## Concluding remarks

The 250 bp 16S mini-barcode used in this study allows wild bee identification of all species, except two specimens of the *Melecta* and *Anthidium* genus. Integrative approaches coupling examination of distance trees, multiple alignment and comparison with morphological data allowed us: 1) to provide 204 new 16S mini-barcodes for wild bees belonging to 93 species verified by taxonomists; 2) to identify species complexes and 3) to delineate efficiently species when females were difficult to separate. This opens avenues for the 16S mini-barcode to be used as an efficient and reliable additional marker in the toolkit for anyone relying on molecular technologies for wild bees ecological studies.

## Acknowledgements

Financial support of this study was provided by ZA-PYGAR 2019-2020. We would like to thank Mathilde Bouchard and Laurent Raison for collecting wild bees legs and Jérôme Willm for collecting wild bees legs and preparing the metadata table. Thanks to Thibault Leroy for critical reviewing of the manuscript. We are grateful to the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, https://doi.org/10.15454/1.5572369328961167E12) for providing computing resources.

## Author contributions

Figure design: AM

Specimen sampling, databasing: AO, RR, AP, NE

Writing of manuscript: MP

Bioanalysis and barcode validation: AM, MO, CK, GP, MP

Editing and comments to the manuscript: AM, MO, CK, GP, AV, AP, NE, KT, EL

Molecular experiments: EL, KT, MP

## References


- Alajmi R, Abdel-Gaber R, Alfozana L (2019) Molecular insights of mitochondrial 16S rDNA genes of the native honey bees subspecies *Apis mellifera carnica* and *Apis mellifera jementica* (Hymenoptera: Apidae) in Saudi Arabia. Genetics and Molecular Research 18 (1): 1-16.
- Almeida EA, Danforth BN (2009) Phylogeny of colletid bees (Hymenoptera: Colletidae) inferred from four nuclear genes. Molecular Phylogenetics and Evolution 50 (2): 290-309. https://doi.org/10.1016/j.ympev.2008.09.028
- Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. Journal of Molecular Biology 215 (3): 403-410. https://doi.org/10.1016/s0022-2836(05)80360-2
- Amiet F (1996) Apidae 1 (*Apis*, *Bombus*, *Psithyrus*). In: CSCF (Ed.) Insecta Helvetica Fauna. 12. Neuchâtel.
- Amiet F (1999) Apidae 2. *Colletes*, *Dufourea*, *Hylaeus*, *Nomia*, *Nomioides*, *Rhophitoides*, *Rophites*, *Sphecodes*, *Systropha*. In: faune CSdcdl (Ed.) Fauna Helvetica. 4. Neuchâtel. URL: https://cscf.abacuscity.ch/de/chf/A~11FHE04X/0~0~Coll/Apidae-2
- Amiet F, Herrmann M, Müller A, Neumeyer R (2001) Apidae 3. *Halictus*, *Lasioglossum*. In: faune CSdcdl (Ed.) Fauna Helvetica. Neuchâtel.
- Amiet F, Herrmann M, Müller A, Neumeyer R (2004) Apidae 4: *Anthidium*, *Chelostoma*, *Coelioxys*, *Dioxys*, *Heriades*, *Lithurgus*, *Megachile*, *Osmia*, *Stelis*. In: CSCF, SEG (Eds) Fauna Helvetica. 9. Neuchâtel.
- Amiet F, Herrmann M, Müller A, Neumeyer R (2007) Apidae 5: *Ammobates*, *Ammobatoides*, *Anthophora*, *Biastes*, *Ceratina*, *Dasypoda*, *Epeoloides*, *Epeolus*, *Eucera*, *Macropis*, *Melecta*, *Melitta*, *Nomada*, *Pasites*, *Tetralonia*, *Thyreus*, *Xylocopa*. In: CSCF, SEG (Eds) Fauna Helvetica. 20. Neuchâtel. URL: https://cscf.abacuscity.ch/de/chf/A~11FHE20X/0~0~Coll/Apidae-5
- Amiet F, Herrmann M, Müller A, Neumeyer R (2010) Apidae 6: *Andrena*, *Melitturga*, *Panurginus*, *Panurgus*. In: CSCF, SEG (Eds) Fauna Helvetica. 26. Neuchâtel. URL: https://cscf.abacuscity.ch/de/chf/A~11FHE26D/0~0~Coll/Apidae-6
- Amiet F, Mueller A, Praz C (2017) Apidae 1: Allgemeiner Teil, Gattungen, *Apis*, *Bombus*. Fauna Helvetica 29: 1-185.
- Aubert M (2020) Proposition de clé d'identification des Eucerini (Hymenoptera: Anthophila) de France continentale – version provisoire. Observatoire des Abeilles. URL: https://oabeilles.net/projets/eucerini
- Avalos G, Trott R, Ballas J, Lin CH, Raines C, Iwanowicz D, Goodell K, Richardson RT (2024) Prospects of pollinator community surveillance using terrestrial environmental DNA metagenetics. Environmental DNA 6: 492. https://doi.org/10.1002/edn3.492

- Benoist R (1931) Les osmies de la faune française. Annales de la Société Entomologique de France 100 (1/2): 23-60. https://doi.org/10.1080/21686351.1931.12280120
- Benoist R (1941) Remarques sur quelques espèces de Mégachiles principalement de la faune française (Hymen Apidae). Annales de la Société entomologique de France 109: 41-88. https://doi.org/10.1080/21686351.1940.12279433
- Biesmeijer JC, Roberts SPM, Reemer M, Ohlemüller R, Edwards M, Peeters T, Schaffers AP, Potts SG, Kleukers R, Thomas CD, Settele J, Kunin WE (2006) Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science 313 (5785): 351-354. https://doi.org/10.1126/science.1127863
- Branstetter MG, Müller A, Griswold TL, Orr MC, Zhu CD (2021) Ultraconserved element phylogenomics and biogeography of the agriculturally important mason bee subgenus Osmia (Osmia). Systematic Entomology 46 (2): 453-472. https://doi.org/10.1111/syen.12470
- Cameron SA, Hines HM, Williams PH (2007) A comprehensive phylogeny of the bumble bees (*Bombus*). Biological Journal of the Linnean Society 91 (1): 161-188. https://doi.org/10.1111/j.1095-8312.2007.00784.x
- Carrié R, Lopes M, Ouin A, Andrieu E (2018) Bee diversity in crop fields is influenced by remotely-sensed nesting resources in surrounding permanent grasslands. Ecological Indicators 90 https://doi.org/10.1016/j.ecolind.2018.03.054
- Casquet J, Thebaud C, Gillespie R (2012) Chelex without boiling, a rapid and easy technique to obtain stable amplifiable DNA from small amounts of ethanol-stored spiders. Molecular Ecology Resources 12 (1): 136-141. https://doi.org/10.1111/j.1755-0998.2011.03073.x
- Cejas D, López-López A, Muñoz I, Ornosa C, De la Rúa P (2019) Unveiling introgression in bumblebee (*Bombus terrestris*) populations through mitogenome-based markers. Animal Genetics 51: 10-1111. https://doi.org/10.1111/age.12874
- Chua PS, Bourlat S, Ferguson C, Korlevic P, Zhao L, Ekrem T, Meier R, Lawniczak MN (2023) Future of DNA-based insect monitoring. Trends in Genetics 39 (7): 531-544. https://doi.org/10.1016/j.tig.2023.02.012
- Clarke L, Soubrier J, Weyrich L, Cooper A (2014) Environmental metabarcodes for insects: *in silico* PCR reveals potential for taxonomic bias. Molecular Ecology Resources 14 (6): 1160-1170. https://doi.org/10.1111/1755-0998.12265
- Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. Nucleic Acids Research 44 (D1): 67-72. https://doi.org/10.1093/nar/gkv1276
- Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Research 16 (22): 10881-10890. https://doi.org/10.1093/nar/16.22.10881
- Costa M, Del Lama M, Melo GR, Sheppard W (2003) Molecular phylogeny of the stingless bees (Apidae, Apinae, Meliponini) inferred from mitochondrial 16S rDNA sequences. Apidologie 34 (1): 73-84. https://doi.org/10.1051/apido:2002051
- Creedy T, Norman H, Tang C, Chin KQ, Andujar C, Arribas P, O'Connor R, Carvell C, Notton D, Vogler A (2020) A validated workflow for rapid taxonomic assignment and monitoring of a national fauna of bees (Apiformes) using high throughput barcoding. bioRxiv https://doi.org/10.1101/575308
- Danforth B (1999) Phylogeny of the bee genus *Lasioglossum* (Hymenoptera: Halictidae) based on mitochondrial COI sequence data. Systematic Entomology 24: 377-393. https://doi.org/10.1046/j.1365-3113.1999.00087.x

- de Carvalho M, Bockmann F, Amorim D, Brandão C, de Vivo M, de Figueiredo J, Britski H, de Pinna MC, Menezes N, Marques FL, Papavero N, Cancello E, Crisci J, McEachran J, Schelly R, Lundberg J, Gill A, Britz R, Wheeler Q, Stiassny MJ, Parenti L, Page L, Wheeler W, Faivovich J, Vari R, Grande L, Humphries C, DeSalle R, Ebach M, Nelson G (2007) Taxonomic impediment or impediment to taxonomy? A commentary on systematics and the cybertaxonomic-automation paradigm. Evolutionary Biology 34: 140-143. https://doi.org/10.1007/s11692-007-9011-6
- Dorchin A, Danforth B, Griswold T (2018) A new genus of eucerine bees endemic to southwestern North America revealed in phylogenetic analyses of the *Eucera* complex (Hymenoptera: Apidae: Eucerini. Arthropod Systematics and Phylogeny 76: 10-3897.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32 (5): 1792-1797. https://doi.org/10.1093/nar/gkh340
- Elbrecht V, Taberlet P, Dejean T, Valentini A, Usseglio-Polatera P, Beisel J, Coissac E, Boyer F, Leese F (2016) Testing the potential of a ribosomal 16S marker for DNA metabarcoding of insects. PeerJ 4 https://doi.org/10.7717/peerj.1966
- Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, Maman S, Hernandez-Raquet G, Combes S, Pascal G (2017) FROGS: Find, Rapidly, OTUs with Galaxy Solution. Bioinformatics 34 (8): 1287-1294. https://doi.org/10.1093/bioinformatics/btx791
- Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3: 294-299.
- Gibbs J, Brady SG, Kanda K, Danforth BN (2012) Phylogeny of halictine bees supports a shared origin of eusociality for *Halictus* and *Lasioglossum* (Apoidea: Anthophila: Halictidae. Molecular Phylogenetics and Evolution 65 (3): 926-939. https://doi.org/10.1016/j.ympev.2012.08.013
- Gibbs J (2018) DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. Genome 61 (1): 21-31. https://doi.org/10.1139/gen-2017-0096
- Hajibabaei M, Alex S, Janzen D, Rodriguez J, Whitfield J, Hebert PN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6 (4): 959-964. https://doi.org/10.1111/j.1471-8286.2006.01470.x
- Herrera-Mesías F, Bause C, Ogan S, Burger H, Ayasse M, Weigand A, Eltz T (2022) Double-blind validation of alternative wild bee identification techniques: DNA metabarcoding and in vivo determination in the field. Journal of Hymenoptera Research 93: 189-214. https://doi.org/10.3897/jhr.93.86723
- Hsieh C, Huang C, Wu W, Wang H (2019) A rapid insect species identification system using mini-barcode pyrosequencing. Pest Management Science 76 (4): 1222-1227. https://doi.org/10.1002/ps.5674
- Janko Š, Rok Š, Blaž K, Danilo B, Andrej G, Denis K, Klemen Č, Matjaž G (2024) DNA barcoding insufficiently identifies European wild bees (Hymenoptera, Anthophila) due to undefined species diversity, genus-specific barcoding gaps and database errors. Molecular Ecology Resources 24 (5). https://doi.org/10.1111/1755-0998.13953
- Kek SP, Chin N, Tan SW, Yusof YA, Chua LS (2017) Molecular identification of honey entomological origin based on bee mitochondrial 16S rRNA and COI gene sequences. Food Control 78: 10-1016.
- Le Divelec R (2021) The West Palaearctic Epeolini Linsley & Michener, 1939 housed in the Muséum national d'Histoire naturelle (Paris) with some taxonomic notes

(Hymenoptera: Apidae: Nomadinae). Annales de la Société entomologique de France N.S. (57): 1-33. https://doi.org/10.1080/00379271.2021.1942206

- Levesque-Beaudin V, Miller M, Dikow T, Miller S, Prosser S, Zakharov E, McKeown J, Sones J, Redmond N, Coddington J, Santos B, Bird J, deWaard J (2023) A workflow for expanding DNA barcode reference libraries through 'museum harvesting' of natural history collections. Biodiversity Data Journal 11 https://doi.org/10.3897/bdj.11.e100677
- Little D (2014) A DNA mini-barcode for land plants. Molecular Ecology Resources 14 (3): 437-446. https://doi.org/10.1111/1755-0998.12194
- Liu M, Clarke L, Baker S, Jordan G, Burridge C (2019) A practical guide to DNA metabarcoding for entomological ecologists. Ecological Entomology 45 (3): 373-385. https://doi.org/10.1111/een.12831
- Magnacca K, Brown MF (2012) DNA barcoding a regional fauna: Irish solitary bees. Molecular Ecology Resources 12 (6): 990-998. https://doi.org/10.1111/1755-0998.12001
- Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2014) Swarm: robust and fast clustering method for amplicon-based studies. PeerJ https://doi.org/10.7717/peerj.593
- Marconi M, Ushiñahua Ramírez D, Cerna Mendoza A, Vecco-Giove CD, Ormeño Luna J, Baikova L, Di Giulio A, Mancini E (2023) An updated molecular phylogeny of the stingless bees of the genus *Trigona* (Hymenoptera, Meliponini) of the northern Peruvian forests. Journal of Hymenoptera Research 96: 751-760. https://doi.org/10.3897/jhr.96.105311
- Marquina D, Andersson A, Ronquist F (2018) New mitochondrial primers for metabarcoding of insects, designed and evaluated using in silico methods. Molecular Ecology Resources 19 (1): 90-104. https://doi.org/10.1111/1755-0998.12942
- Michener CD (2007) The bees of the world. Johns Hopkins University Press https://doi.org/10.56021/9780801885730
- Müller A (2024) Palaearctic Osmiine Bees, ETH Zürich. http://blogs.ethz.ch/osmiini
- Newton J, Bateman P, Heydenrych M, Kestel J, Dixon K, Prendergast K, White N, Nevill P (2023) Monitoring the birds and the bees: Environmental DNA metabarcoding of flowers detects plant–animal interactions. Environmental DNA 5 (3): 488-502. https://doi.org/10.1002/edn3.399
- Odanaka K, Branstetter M, Tobin K, Rehan S (2022) Phylogenomics and historical biogeography of the cleptoparasitic bee genus *Nomada* (Hymenoptera: Apidae) using ultraconserved elements. Molecular Phylogenetics and Evolution 170 https://doi.org/10.1016/j.ympev.2022.107453
- Ollivier M, Rivers-Moore J, Pichon M, Andrieu E, Carrié R, Rudelle R, Sarthou J, Ouin A (2024) Wild bees (Apoidea, Anthophila) of south-west France: more than 10 years of inventories in mosaic landscapes of "Vallées et Coteaux de Gascogne" (ZA-PYGAR). Biodiversity Data Journal https://doi.org/10.3897/BDJ.12.e135157
- Ollivier M, Rudelle R, Genoud D, Marquisseau A, Rougerie R, Perrard A, Pichon M (In prep.) A national reference CO1 DNA barcode database covering 80% of the French wild bees fauna (*Apoidea antophila*). BMC Ecology & Evolution.
- Ornosa C, Ortiz-Sanchez F (2004) Hymenoptera: Apoidea I. In: Ramos MA, et al. (Ed.) Fauna Ibérica. 23. Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Cientificas, Madrid.
- Ortega A, Geraldi N, Díaz-Rúa R, Ørberg S, Wesselmann M, Krause-Jensen D, Duarte C (2020) A DNA mini-barcode for marine macrophytes. Molecular Ecology Resources 20 (4): 920-935. https://doi.org/10.1111/1755-0998.13164

- Ortiz-Sanchez FJ, Jimenez-Rodriguez AJ (1991) Actualizacion del catalogos de las especies españolas de Anthophorini (Hymenoptera, Anthophoridae). Boletin de la Asociacion española de Entomologia.
- Ouin A, Andrieu E, Vialatte A, Balent G, Barbaro L, Blanco J, Ceschia E, Clement F, Fauvel M, Gallai N, Hewison AJM, Jean-François D, Kephaliacos C, Macary F, Probst A, Probst J, Ryschawy J, Sheeren D, Sourdril A, Tallec T, Verheyden H, Sirami C (2021) Chapter Two - Building a shared vision of the future for multifunctional agricultural landscapes. Lessons from a long term socio-ecological research site in south-western France. In: Bohan D, Dumbrell A, Vanbergen A (Eds) Advances in Ecological Research. Vol. 65. Academic Press https://doi.org/10.1016/bs.aecr.2021.05.001
- Packer L, Ruz L (2017) DNA barcoding the bees (Hymenoptera: Apoidea) of Chile: species discovery in a reasonably well known bee fauna with the description of a new species of *Lonchopria* (Colletidae). Genome 60 (5): 414-430. https://doi.org/10.1139/gen-2016-0071
- Pauly A (2015) Clé illustrée pour l'identification des abeilles de Belgique. II. Megachilidae. Document de travail du projet BELBEES.
- Pauly A (2019) Abeilles de Belgique et des régions limitrophes (Insecta Apoidea). Famille Halictidae. Institut royal des Sciences naturelles de Belgique
- Pauly A, Noël G, Sonet G, Notton D, Boevé J-L (2019) Integrative taxonomy resuscitates two species in the *Lasioglossum villosulum* complex (Kirby, 1802) (Hymenoptera: Apoidea: Halictidae). European Journal of Taxonomy 541 https://doi.org/10.5852/ejt.2019.541
- Piper AM, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, Blacket MJ (2019) Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. GigaScience 8 (8). https://doi.org/10.1093/gigascience/giz092
- Powney G, Carvell C, Edwards M, Morris RA, Roy H, Woodcock B, Isaac NB (2019) Widespread losses of pollinating insects in Britain. Nature Communications 10 (1). https://doi.org/10.1038/s41467-019-08974-9
- Praz C, Müller A, Danforth B, Griswold T, Widmer A, Dorn S (2008) Phylogeny and biogeography of bees of the tribe Osmiini (Hymenoptera: Megachilidae). Molecular Phylogenetics and Evolution 49: 185-197. https://doi.org/10.1016/j.ympev.2008.07.005
- Rasmont P (2014) Clef des Melectini de France (Apidae). Université de Mons, Belgique.
- Rasmont P, Ghisbain G, Terzo M (2021) Bumblebees of Europe and neighbouring regions. NAP Editions, Verrières-le-Buisson.
- Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular ecology notes 7 (3): 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x
- Ratnasingham S, Hebert PN (2013) A DNA-based registry for all animal species: The Barcode Index Number (BIN) System. PLOS One 8 (7). https://doi.org/10.1371/journal.pone.0066213
- Ratnasingham S, Wei C, Chan D, Agda J, Agda J, Ballesteros-Mejia L, Boutou HA, El Bastami ZM, Ma E, Manjunath R, Rea D, Ho C, Telfer A, McKeowan J, Rahulan M, Steinke C, Dorsheimer J, Milton M, Hebert PD (2024) BOLD v4: A centralized bioinformatics platform for DNA-based biodiversity data. Methods in Molecular Biology 2744: 403-441. https://doi.org/10.1007/978-1-0716-3581-0_26
- Remmel N, Buchner D, Enss J, Hartung V, Leese F, Welti EA (2024) DNA metabarcoding and morphological identification reveal similar richness, taxonomic composition and body

size patterns among flying insect communities. Insect Conservation and Diversity 17 (3): 449-463. https://doi.org/10.1111/icad.12710

- Rhodes C (2018) Pollinator decline – An ecological calamity in the making? Science Progress 101 (2): 121-160. https://doi.org/10.3184/003685018x15202512854527
- Rivers-Moore J, Andrieu E, Vialatte A, Ouin A (2020) Wooded semi-natural habitats complement permanent grasslands in supporting wild bee diversity in agricultural landscapes. Insects 11: 812-10. https://doi.org/10.3390/insects11110812
- Santos B, Miller M, Miklasevskaja M, McKeown J, Redmond N, Coddington J, Bird J, Miller S, Smith A, Brady S, Buffington M, Chamorro ML, Dikow T, Gates M, Goldstein P, Konstantinov A, Kula R, Silverson N, Solis MA, deWaard S, Naik S, Nikolova N, Pentinsaari M, Prosser S, Sones J, Zakharov E, deWaard J (2023) Enhancing DNA barcode reference libraries by harvesting terrestrial arthropods at the Smithsonian's National Museum of Natural History. Biodiversity Data Journal 11 https://doi.org/10.3897/bdj.11.e100904
- Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau D, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick B, Pruitt K, Sherry S (2021) Database resources of the national center for biotechnology information. Nucleic Acids Research 50 https://doi.org/10.1093/nar/gkab1112
- Schmid-Egger C, Scheuchl E (1997) Illustrierte Bestimmungstabellen der Wildbienen Deutschlands und sterreichs 3: Andrenidae. Self-publishing
- Schmidt S, Schmid-Egger C, Morinière J, Haszprunar G, Hebert PN (2015) DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). Molecular Ecology Resources 15 (4): 985-1000. https://doi.org/10.1111/1755-0998.12363
- Sheffield C, Hebert PN, Kevan P, Packer L (2009) DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. Molecular Ecology Resources 9: 196-207. https://doi.org/10.1111/j.1755-0998.2009.02645.x
- Sheffield C, Heron J, Gibbs J, Onuferko T, Oram R, Best L, deSilva N, Dumesh S, Pindar A, Rowe G (2017) Contribution of DNA barcoding to the study of the bees (Hymenoptera: Apoidea) of Canada: progress to date. The Canadian Entomologist 149 (6): 736-754. https://doi.org/10.4039/tce.2017.49
- Sickel W, Kulow J, Krüger L, Dieker P (2023) BEE-quest of the nest: A novel method for eDNA-based, nonlethal detection of cavity-nesting hymenopterans and other arthropods. Environmental DNA 5: 1163-1176. https://doi.org/10.1002/edn3.490
- Sless TJ, Branstetter MG, Mikát M, Odanaka KA, Tobin KB, Rehan SM (2024) Phylogenomics and biogeography of the small carpenter bees (Apidae: Xylocopinae: Ceratina). Molecular Phylogenetics and Evolution 198 https://doi.org/10.1016/j.ympev.2024.108133
- Smit J (2018) Identification Key to the European Species of the Bee Genus Nomada Scopoli, 1770 (Hymenoptera: Apidae), Including 23 New Species. Museum Witt
- Straka J, Benda D, Policarová J, Astapenková A, Wood T, Bossert S (2024) A phylogenomic monograph of West-Palearctic *Nomada* (Hymenoptera: Apidae). Insect Systematics and Diversity 8: 10-1093.
- Sun C, Huang J, Wang Y, Zhao X, Su L, Thomas GC, Zhao M, Zhang X, Jungreis I, Kellis M, Vicario S, Sharakhov I, Bondarenko S, Hasselmann M, Kim C, Paten B, Penso-Dolfin L, Wang L, Chang Y, Gao Q, Ma L, Ma L, Zhang Z, Zhang H, Zhang H, Ruzzante L,

Robertson H, Zhu Y, Liu Y, Yang H, Ding L, Wang Q, Ma D, Xu W, Liang C, Itgen M, Mee L, Cao G, Zhang Z, Sadd BM, Hahn MW, Schaack S, Barribeau SM, Williams PH, Waterhouse RM, Lockridge Mueller R (2021) Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. Molecular Biology and Evolution 38 (2): 486-501. https://doi.org/10.1093/molbev/msaa240

- Terzo M, Iserbyt S, Rasmont P (2007) Révision des Xylocopinae (Hymenoptera : Apidae) de France et de Belgique, Annales de la Société entomologique de France. International Journal of Entomology 43 (4): 445-491. https://doi.org/10.1080/00379271.2007.10697537
- Trianto M, Purwanto H (2020) Molecular phylogeny of stingless bees in the special region of Yogyakarta revealed using partial 16S rRNA mitochondrial gene. Buletin Peternakan 44 (4). https://doi.org/10.21059/buletinpeternak.v44i4.55539
- Villalta I, Ledet R, Baude M, Genoud D, Bouget C, Cornillon M, Moreau S, Courtial B, Lopez-Vaamonde C (2021) A DNA barcode-based survey of wild urban bees in the Loire Valley, France. Scientific reports 11 (1): 4770. https://doi.org/10.1038/s41598-021-83631-0
- Vinarski M (2020) Roots of the taxonomic impediment: Is the “integrativeness” a remedy? Integrative Zoology 15 (1): 2-15. https://doi.org/10.1111/1749-4877.12393
- Whitfield JB, Cameron SA (1998) Hierarchical analysis of variation in the mitochondrial 16S rRNA gene among Hymenoptera. Molecular Biology and Evolution 15 (ue 12): 1728-1743. https://doi.org/10.1093/oxfordjournals.molbev.a025899
- Wood T, Ghisbain G, Michez D, Praz C (2021) Revisions to the faunas of *Andrena* of the Iberian Peninsula and Morocco with the descriptions of four new species The InBIO Barcoding Initiative Database: DNA barcodes of Iberian Bees 69 (Hymenoptera: Andrenidae. European Journal of Taxonomy 758: 147-193. https://doi.org/10.5852/ejt.2021.758.1431
- Wood T (2023) The genus Andrena Fabricius, 1775 in the Iberian Peninsula (Hymenoptera, Andrenidae). Journal of Hymenoptera Research 96: 241-484. https://doi.org/10.3897/jhr.96.101873
- Wood T, Gaspar H, Le Divelec R, Penado A, Silva TL, Mata V, Veríssimo J, Michez D, Castro S, Loureiro J, Beja P, Ferreira S (2024) The InBIO Barcoding initiative database: DNA barcodes of Iberian bees. Biodiversity Data Journal 12 https://doi.org/10.3897/bdj.12.e117172
- Yoon HJ, Kim SE, Kim YS, Lee SB (2004) Colony developmental characteristics of the bumblebee queen *Bombus ignitus* by the first oviposition day. International Journal of Industrial Entomology 8: 139-143.

## Supplementary materials

### Suppl. material 1: 16S Wild Bees mini-barcodes from Occitania: metadata doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon

**Data type:** Metadata

Download file (166.08 kb)

## Suppl. material 2: 16S Wild Bees mini-barcodes from Occitania: DNA sequences doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon
**Data type:** DNA Sequences, Fasta
Download file (97.64 kb)

## Suppl. material 3: Lost species doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon
**Data type:** List
**Brief description:** List of the 23 species eliminated during the barcode acquisition process.
Download file (1.23 kb)

## Suppl. material 4: Available 16S in GenBank doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon
**Data type:** List, Accession numbers
Download file (5.70 kb)

## Suppl. material 5: Sequencing and barcoding success per species doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon
**Data type:** Occurrences
Download file (5.86 kb)

## Suppl. material 6: Table of intraspecific distances doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon
**Data type:** Genetic distances
**Brief description:** Distances between specimens within species. To find minimum and maximum intraspecific distance for a specific species, filter the column "Species".
Download file (18.42 kb)

## Suppl. material 7: Table of interspecific distances doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon

**Data type:** Genetic distances

**Brief description:** Distances between specimens belonging to different species within their genus. To find minimum and maximum interspecific distance between two species, filter the columns "species_1" and "species_2". To find minimum and maximum intragenus distance, filter the column "Genus".

Download file (803.68 kb)

## Suppl. material 8: Andrenidae Tree doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon

**Data type:** NJ Tree

Download file (99.87 kb)

## Suppl. material 9: Apidae Tree doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon

**Data type:** NJ Tree

Download file (116.09 kb)

## Suppl. material 10: Colletidae Tree doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon

**Data type:** NJ Tree

Download file (86.46 kb)

## Suppl. material 11: Megachilidae Tree doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon

**Data type:** NJ Tree

Download file (89.08 kb)

## Suppl. material 12: Halictidae Tree doi

**Authors:** Anaïs Marquisseau, Kamila Canale-Tabet, Emmanuelle Labarthe, Géraldine Pascal, Christophe Klopp, André Pornon, Nathalie Escaravage, Rémi Rudelle, Alain Vignal, Annie Ouin, Mélodie Ollivier, Magalie Pichon

**Data type:** NJ Tree

Download file (95.15 kb)